Domain Wise Distribution of Mutations in Dystrophin Protein and Duchenne Muscular Dystrophy

Simanti Bhattacharya1, Amit Das1, Angshuman Bagchi* 
Department of Biochemistry and Biophysics, University of Kalyani, Nadia-741235, West Bengal, India

Abstract
Duchenne muscular dystrophy, the most common inherited X-linked recessive muscular dystrophy, affects 20000 new borns per year globally. Since its discovery in 1860, extensive research works have been carried out to understand the complex architecture of the disease formation. Reason behind the onset of the disease has been mapped back to the set of mutations of different types in dystrophin gene (DMD, 2.4 million bp), the largest gene in the body. Dystrophin (Dp), the cytosolic protein acts as the root of the complex which was primarily thought to link extracellular matrix with cellular actin cytoskeleton but later on has been associated with the stability of the cells, signal transduction as well as in proper development. In this review, we gathered the details of all the mutations occurring in DMD gene and observed that majority of the mutations are present in the N terminal Actin Binding Domain. Some of the mutations were found to be present in the Cysteine Rich Domain of the protein, reflecting the point that these two domains are the most mutation prone regions contributing to Duchenne Muscular Dystrophy (DMD) onset. This review therefore gives an integrative view describing the involvements of Dp in regulating the complexity of DMD disease with a future aspect to study the structural details of DMD genes along with its genetic variations.

Keywords: Duchenne muscular dystrophy; Dystrophin; Dystrophin associated protein complex; Actin cytoskeleton; Mutations; Therapeutics

Abbreviations: Dp: Dystrophin; DMD: Duchenne Muscular Dystrophy; DG: Dystroglycan; α-DG: Alpha Dystroglycan; β-DG: Beta Dystroglycan; bp: Base Pair; CNS: Cerebro Nervous System; DAPC: Dystrophin Associated Protein Complex; SGC: Sarcoglycans; SS: Sarcospan; BMD: Becker Muscular Dystrophy; DCM: Dilated Cardiomyopathy; AON: Antisense oligonucleotides

Introduction
The mechanical strength of a cell is provided by the mesh network of cytoskeletal proteins: thin filaments (actin cytoskeleton), intermediate filaments and thick filaments (microtubules; tubulin) [1]. Unique to its nature, actin filament, among them, possesses continuous tread milling nature [2]. Actin cytoskeleton, thought initially was thought to act as scaffold protein to provide mechanical strength to a cell, is now very much associated with endocytosis, exocytosis, cell polarity, cell movement, signal transduction as well as it has been proven to be essential for growth and development along with playing important role in neuronal tube closure etc. [3,4]. In muscle cell membrane, cellular actin cytoskeleton is linked to extracellular matrices via a large macromolecular organization of protein complex known as Dystrophin Associated Protein Complex (DAPC) [5] which has cytosolic protein Dystrophin (Dp) [6] establishing the direct interaction with actin cytoskeleton. The discovery of Dp protein is associated with the identification of the causative agent of Duchenne Muscular Dystrophy (DMD), the most common lethal X-linked muscular disease causing progressive muscle tissue degeneration with early death of patients [7]. DMD gene has been found to be mutated at several points with missense point mutation, nonsense point mutations, deletions, duplications, frame shift etc, either producing functionally inactive protein product (Dp malfunctioning) or ending up with abortive translation (Dp deficiency). Every year about 20,000 children are born with this disease globally [8]. Several groups have classified these mutations and have also shown which types of mutations occur majorly in DMD patients [9-11]. In our review we have described Dp and its associated partner protein complex as well as have given a glance to the severity of DMD and possible therapeutic approaches, direct or secondary, available till date [12,13]. We have focused majorly to identify which domain(s) is/are more susceptible to mutations and to which type of mutations. And finally our review draws readers’ attention to detect how these mutations in those susceptible domains can destabilize the bridge linking actin cytoskeleton and extra cellular matrix.

Dystrophin Protein and Associated Protein Complex

Dystrophin
Dystrophin (Dp, 427 kDa) is a cytosolic protein, widely expressed in varied types of tissues and works as the major protein to complete the link between extracellular matrix and cytosolic actin cytoskeleton [14]. This protein was originally identified by Louis M Kunkel and his group in the year of 1986 while working with patients suffering from Duchenne Muscular Dystrophy (DMD) [7]. Dp is the product of the largest gene known till date, the DMD gene. This gene, in other sense carries a classic example of complex regulation of genetic transcription and takes approximately 16 hours to be transcribed [15]. It contains 79 exons interspaced by introns of varied length. The major promoters for expression of the full length protein product are located in the 5’ region of the gene which is generally present 320 kb upstream of exon 2 (Figure 1A). Major isoforms of Dp includes Dp427, Dp260, Dp140, Dp116, Dp71 and Dp40 [16]. Isoforms are named after their molecular weight as indicated by the number suffixed to ”Dp” (Dp427 or Dp116), their tissue specific promoter mediated transcript (Dp427m for muscle or Dp427c for cortical) and by alternative spliced products (Dp71b or Dp71c) [14]. Dp427 is the major muscle isoform of Dp and has been considered to be the canonical form [17]. Dystrophin protein

*Corresponding author: Angshuman Bagchi, Department of Biochemistry and Biophysics, University of Kalyani, Nadia-741235, West Bengal, India, Tel: +91-9051948843; E-mail: angshu@klyuniv.ac.in

Received September 01, 2015; Accepted September 15, 2015; Published September 18, 2015


Copyright: © 2015 Bhattacharya S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
is a multi-domain protein consisting of four distinguished domains with specified biological functions (Figure 1B). Starting from amino terminus end the domains are:

(1) Amino terminal Actin Binding Domain (ABD, 1-246 amino acid residues) that contains two calponin homology domain which directly interacts with cellular actin cytoskeleton [18].

(2) A central rod shaped domain consisting 24 spectrin like repeats (SpR, 339-3040 amino acid residues). Proline rich hinge region separates the Spectrin like repeats. This rod domain with hinge has been thought to provide the stretching and flexibility of the protein to regulate actin dynamical force properly [19].

(3) Cysteine Rich Domain (CRD, 3055-3360 amino acid residues) is a collection of small functionally active domains- WW domain, Ca2+ dependent EF hands and ZZ domains. This CRD domain establishes the interaction between beta dystroglycan (β-DG) and Dp [20].

(4) And finally there is the fourth domain the C terminal region (C-Term, 3361-3865 amino acid residues). Dp bears high similarity with other known actin binding proteins for the rest of its domains [21]. But the C terminus of Dp is unique as this domain carries out major interactions with downstream partner proteins like Dystrobrevin and Syntrophins [22].

Isosforms of Dp generally differ from canonical form at their N terminal region. Dp260 [23-24] is found in retina and is very crucial to normal retinal electrophysiology [23]. The presence of 13 novel amino acids as its N-terminal makes it different from its canonical isoform and this unique N terminus is followed by most of the SpR, CRD and C-Term. Dp140 [25,26] is expressed throughout the CNS and is transcribed by an alternative promoter residing upstream to exon 45 in the dystrophin locus. This also has altered N terminal region. Transcription of Dp116 [27] is initiated at an exon located approximately 850 bp upstream of DMD gene exon 56 and is majorly expressed in adult peripheral nerves. Dp71 [28] is found in brain and other non-muscle tissues. Discovery of this 71kDa isoform of Dp suggested the existence of another promoter situated upstream of exon 63 [28]. It does not have an intact putative WW domain. But we have found that despite the presence of its truncated WW domain, it can efficiently interact with poly proline rich region of β-DG [29]. Dp71 is not expressed in skeletal muscle tissues but is the major product in non-muscle tissues. Its immense importance reflects with the notation that this is present in embryonic stem cells and is the first gene product to be identified in developmental stage [30]. In our earlier works, we have gathered information from publicly available DMD database sources and have isolated 18 novel point mutation causing DMD in particular [31]. Majority of these point mutations are found to be localized at the Cysteine Rich Region (CRD). These point mutations do not cause abortion of Dp protein expression but are largely associated with malfunctioning protein formation. Analyzing the effects of these point mutations individually on the protein structure, di-sulphide bond, hydrogen bond formation, accessible surface area we have found that the intact architecture of actin binding site at ABD is disturbed, changes in accessible surface area, secondary structures and so on. These helped us to understand the possible mechanism lying behind the weak interaction of Dp with its partner proteins in presence of the mutation. Moreover, most severe effect is caused when a Cysteine (Cys) residue is replaced by Arginine (Arg) or any other amino acid [31].

**Dystrophin Associated Protein Complex (DAPC)**

Discovery of Dp as the causative agent behind DMD, tempted researchers to look into the associated partner proteins. Several experiments, carried in this field, revealed the detailed protein complex associated with Dp in normal tissue [32] as well as in dystrophic tissues [33]. All these large macromolecular complex of proteins are collectively coined as Dystrophin Associated Protein Complex (DAPC) [5]. Frontier members of this complex (Figure 2) are Dystroglycan (DG), a widely expressed glycoprotein with alpha subunit (α-DG residing...
at the extracellular surface of cells and beta subunit (β-DG), which is the membrane spanning part; Sarcoglycans (SGC), a five membered transmembrane protein; Sarcospan (SS), a 25 kDa membrane protein possessing four trans-membrane domains with its N- and C-termini and is located intra-cellular; Alpha Dystrobrevin containing two tandem alpha helical syntrophin binding sites and several tyrosine kinase consensus sites and finally Syntrophins, containing PDZ domain capable of facilitating homo- and heterodimerization with other PDZ-containing proteins.

Alpha Dystroglycan (α-DG) receives signal from extracellular matrix proteins like laminin, Parlecain, Agrin etc and transmits the signals to its membrane bound counterpart β-DG. Proper glycosylation of α-DG at its mucin rich region is necessary for this interaction [5,34]. Our work [34] on understanding the effect of a naturally causing mutation T192M onto α-DG structure and interaction with its ligand Laminin, an extra cellular matrix protein have revealed that the replacement of Threonine (Thr, T) with Methionine (Met, M) has brought out surface hydrophobicity changes and compromised intra protein hydrogen bonds weakening the protein stability. Furthermore, studies following MD simulation have shown that in the presence of mutation the Cys182-Cys264 S-S bond, crucial to maintain the N-terminal globular domain architecture, is disturbed [35]. These findings guided us to understand the reason behind the weak interaction of α-DG with Laminin and also helped us to explore the mechanism behind the onset of Muscle Dysrophy, Dystroglycanopathy, Type C, 9 [OMIM 613818]. The β-DG, on the other hand, directly interacts with CRD of Dp which in turn interacts directly with cellular actin cytoskeleton and thereby contributes in signaling cascade [36].

**Mutation in Dystrophin and Muscular Dystrophy**

**Duchenne muscular dystrophy (DMD)**

The large gene size in association with alternative splicing itself is responsible for large number of mutations in DMD gene, encoding Dp protein. Deficiency of a functional Dp leads to a spectrum of complicated diseases called dystrophinopathies [37]. Three major forms of dystrophinopathies are: Dilated Cardio Myopathy [DCM], Becker Muscular Dysrophy [BMD] and Duchenne Muscular Dysrophy [DMD]. Among these three, DMD has been reported to be the most lethal as well as common and incurable form of three dystrophinopathies. It can be found in every 1 in 3500 new borns [38]. French neurologist Guillaume Benjamin Amand Duchenne, in 1860, discovered this disease and described it as lethal X linked inherited disease. DMD has an early onset of symptoms with loss of ambulance at the age of 9-13 years and eventual death [9,38]. With the advancement of modern technologies in scientific research, some approaches have been generated to compensate the diseased condition [39]. Most advanced among them is the use of antisense oligonucleotides (AON) synthesized in such a way to fix exon skipping and to restore the reading frame to synthesize small but functional Dp protein [40]. In another approach, utrophin short segments are being delivered to dystrophic cells as a transgene cloned in non-adeno viral vectors. Small non coding RNAs have shown promising protection against DMD [41]. Prednisolone and its derivatives are also used to prevent the premature death of Duchenne patients as these glucocorticoids confer therapeutic advantages by strengthening muscles [42]. Again stem cells to regenerate muscle tissue have also proven to be promising in therapeutic approaches for DMD [43]. But currently most potential drugs are exon skipping agents eteplersen and drisapersen [44].

**Distribution of Mutations Leading to DMD**

Mutations in DMD gene leads to dystrophinopathic state. In DMD, as mentioned earlier, functional Dp protein is absent majorly. But looking into the molecular genetics has revealed that numbers of different types of mutations have played a combinatorial role to obstruct a functional Dp protein production. In their database reports, Sylvie Tuffery-Giraud et al. have figured out that among all the mutations that occurred in DMD patients, 61% were deletions, 13% were duplications and 26% were point mutations [45]. In fact, several experiments conducted with DMD patients from different origin have also indicated the same distribution pattern of mutation [9,46,47]. Inspired by this finding we have also tried to analyze the distribution of several types of mutations occurring in exons of DMD gene, informations gathered from the samples available in UMD-DMD database (http://www.umd.be/DMD/W_DMD/search.shtml) and also tried to map back the distribution of those mutations in the four major domains of Dp protein. Mutations have been grouped as: missense mutations, nonsense mutations, small lesions (i.e. deletions or insertions <1 exon or spicing sites <10bp from exon) and large lesions (large deletions or duplication spanning >1 exon). The domains have been marked as per their respective exons. The charts (Figure3) demonstrate that major of large duplications or deletions have occurred in the starting exons (exon 2-8) of DMD gene , or alternatively we can say, at the ABD of Dp. In fact all types of possible mutations are occurring in this region and shares a large percentage, establishing it as the highly mutation prone unit. Next in the position comes CRD which harbors majority of missense mutations that may alter protein structure and small deletion or insertions that majorly account for frame shift mutations. Rest two domains show variable percentage of mutational types (Table 1). If we correlate this finding with the above one, documented in ref. [45-47], then we can say ABD and CRD are majorly mutation susceptible domains.

**Biological Implementations**

DMD, the most common form of muscular dystrophies, is associated with muscle deformation and death at early ages. DMD is an inherent X-linked muscular dystrophy generated due to the mutations in DMD gene, encoding Dystrophin protein. As discussed above, not only one specific mutation but rather combinatorial effects...
of deletion, duplications, frame shift mutations in DMD gene decide the complexity and severity of DMD in patients. Extensive works to prevent this deleterious disease have provided several therapeutic approaches like gene editing, stem cell therapy, exon skipping, immunomodulations, stop codon read through for the restoration of ORF coding functional Dp, viral-non viral vectors carrying minigene of utrophin, increased protein thiol oxidation as well as applications of glucocorticoids etc [25,48-52]. But none of these approaches alone can cure the disease completely, majorly because in Dp malfunctioning or deficiency, not only DAPC mediated neuronal signal gets destabilized but also a drastic change occurs in the flux of ion current through the ion channels in cell membrane [44]. It is worth mentioning that most of the therapeutic approaches either directly manipulate the genetic drifts (like exon skipping, gene editing etc) or acquire a secondary by-pass (like glucocorticoids, thiol oxidation etc). Only few approaches exist that involve application of structural homologue of Dp, like construct of mini-utrophin-gene or short constructs of DMD genes [53]. These methods deal with the structural aspects and compensation of malfunctioning Dp. In our review we intended to describe the structural organization of Dp and to identify the mutational distribution of the different domains of Dp protein. And we have successfully marked the ABD and CRD are the hub of the major mutational events occurring. It is worth mentioning here that these two ends of Dp actually are the two major points for maintaining the bridge to link actin cytoskeleton and extracellular matrix. In our review we have also focused on the contribution and interaction pattern of these two domains with their immediate interaction partners; actin for ABD and β-DG for CRD. Moreover this review also describes how mutations in those domains associate a cell’s fate to fatal DMD consequences. To understand a disease mechanism and to prevent it, insight into the structural aspects of a protein carries immense importance as of the genetic detailing. In conclusive remarks we therefore hypothesize that detailed investigations of the effects of the mutations in the structural stability as well as in the interaction pattern will decipher the disease mechanism in a greater detail. Approaches with genetic manipulations or transcription level control are expensive and require targeted delivery with proper control measures. Similarly secondary treatments like drug delivery, immunomutations are associated with post treatment toxic accumulations. But supplements with structural homologues like utrophin C terminus or Dp C terminus constructs in non adenoviral vectors can be proven to be more target specific in curing the disease along with other currently available approaches.

**Conclusion**

DMD gene, the largest gene in the body acquires several types of mutations, most of which lead to abortive translation with no Dystrophin protein production. But certain mutations are there that end up in a malfunctioning Dystrophin protein production. These mutational events in DMD gene, with multiple combination lead to muscular dystrophies, DMD being the most lethal one among them. There remains a huge need to understand the mutations that are associated with malfunctioning protein production for Dystrophin, not only because these mutations alter protein structure but also they hamper the natural protein-protein interaction cascade. This review work has been done to discuss the DMD gene, its protein (Dystrophin), the protein associated complex and the muscular dystrophies occurring due to mutations. Details for mutations, irrespective of the type of mutations, have been collected from available DMD databases which enlist the mutation position in the exon and corresponding position in the amino acid position and the change in amino acids due to the gene level mutations. Only those novel mutations have been collected that are directly linked with the generation of DMD or DMD/BMD diseases. The review thereafter was directed to analyze the domain wise distribution of those mutations. So the review directly focuses into the mutations occurring in the gene and subsequently translated to protein leaving the protein not functioning properly and giving rise to DMD. The review also discussed the available therapeutics developed till date. The review will therefore be valuable to understand the mutations, their effects and distributions and mechanisms of DMD onset.

**Acknowledgements**

The authors are really grateful to the BIF Center, University of Kalyani for providing the necessary equipments and workstation to carry out the review. SB and AD also are thankful to UGC, India and CSIR, India for their respective fellowships. The authors would like to acknowledge the DST-PURSE program 2012-2015 going on in the department of Biochemistry and Biophysics, University of Kalyani and the DBT, India (project no. BT/PRR/6809/BID/7/417/2013) for the support.

**References**


